Cuestionario de teoría 3

Antonio Álvarez Caballero
analca3@correo.ugr.es

6 de junio de 2016

1. Cuestiones

Cuestión 1. Considera los conjuntos de hipótesis $\mathcal{H}_1$ y $\mathcal{H}_{100}$, que contienen funciones Booleanas sobre 10 variables Booleanas, es decir $\mathcal{X} = \{-1, +1\}^{10}$. $\mathcal{H}_1$ contiene todas las funciones Booleanas que toman valor +1 en un único punto de $\mathcal{X}$ y -1 en el resto. $\mathcal{H}_{100}$ contiene todas las funciones Booleanas que toman valor +1 exactamente en 100 puntos de $\mathcal{X}$ y -1 en el resto.

a) ¿Cuántas hipótesis contienen $\mathcal{H}_1$ y $\mathcal{H}_{100}$?

b) ¿Cuántos bits son necesarios para especificar uan de las hipótesis en $\mathcal{H}_1$?

c) ¿Cuántos bits son necesarios para especificar uan de las hipótesis en $\mathcal{H}_{100}$?

Argumente sobre la relación entre la complejidad de una clase de funciones y la complejidad de sus componentes.

Solución. solucion

Cuestión 2. cuestion

Solución. solucion

Cuestión 3. En un experimento para determinar la distribución del tamaño de los peces en un lago, se decide echar una red para capturar una muestra representativa. Así se hace y se obtiene una muestra suficientemente grande de la que se pueden obtener conclusiones estadísticas sobre los peces del lago. Se obtiene la distribución de peces por tamaño y se entregan las conclusiones. Discuta si las conclusiones obtenidas servirán para el objetivo que se persigue e identifique si hay algo que lo impida.

Solución. solucion

Cuestión 4. Considere la siguiente aproximación al aprendizaje. Mirando los datos, parece que los datos son linealmente separables, por tanto decidimos usar un simple perceptrón y obtenemos un error de entrenamiento cero con los pesos óptimos encontrados. Ahora deseamos obtener algunas conclusiones sobre generalización, por tanto miramos el valor d_{VC} de nuestro modelo y vemos que es $d + 1$. Usamos dicho valor de d_{VC} para obtener una cota del error de test. Argumente a favor o en contra de esta forma de proceder identificando los posibles fallos si los hubiera y en su caso cuál hubiera sido la forma correcta de actuación.

Solución. solucion

Cuestión 5. Suponga que separamos 100 ejemplos de un conjunto $\mathcal{D}$ que no serán usados para entrenamiento, sino que serán usados para seleccionar una de las tres hipótesis finales g_1, g_2, g_3 producidas por tres algoritmos de aprendizaje distintos entrenados sobre el resto de datos. Cada algoritmo trabaja con un conjunto $\mathcal{H}$ de tamaño 500. Nuestro deseo es caracterizar la precisión de la estimación $E_{out}(g)$ sobre la hipótesis final seleccionada cuando usamos los mismos 100 ejemplos para hacer la estimación.

a) ¿Qué expresión usaría para calcular la preción? Justifique la decisión.

b) ¿Cuál es el nivel de contaminación de estos 100 ejemplos comparándolo con el caso donde estas muestras fueran usadas en el entrenamiento en lugar de en la selección final?

Solución. solucion

Cuestión 6. Considere la tarea de seleccionar una regla del vecino más cercano. ¿Qué hay de erróneo en la siguiente lógica que se aplica a la selección de k? (Los límites son cuando $N \to \infty$). "Considere la posibilidad de establecer la clase de hipótesis $\mathcal{H}_{NN}$ con N reglas, las k-NN hipótesis, usando $k = 1, \ldots, N$. Use el error dentro de la muestra para elegir un valor de k que minimiza E_{in}. Utilizando el error de generalización para N hipótesis, obtenemos la conclusión de que $E_{in} \to E_{out}$ porque $\frac{\log(N)}{N} \to 0$. Por lo tanto concluimos que asintóticamente, estaremos eligiendo el mejor valor de k, basándonos sólo en E_{in}".

Solución. solucion

Cuestión 7. Responder estas cuestiones:

a) Considere un núcleo Gaussiano en un modelo de base radial. ¿Qué representa $g(x)$ (ecuación 6.2 del libro LfD) cuando $||x|| \to \infty$ para el modelo RBF no-paramétrico vs el modelo RBF paramétrico, asumiendo w_n fijos?

b) Sea $\mathcal{Z}$ una matriz cuadrada de características definida por $\mathcal{Z}_{nj} = \Phi_j(x_n)$, donde $\Phi_j(x)$ representa una transformación no lineal. Suponer que $\mathcal{Z}$ es invertible. Mostrar que un modelo paramétrico de base radial, con $g(x) = w^T\Phi(x)$ y $w = \mathcal{Z}^{-1}y$, interpola los puntos de forma exacta. Es decir, que $g(x_n) = y_n$, con $E_{in}(g) = 0$.

c) ¿Se verifica siempre que $E_{in}(g) = 0$ en el modelo no paramétrico?

Solución. solucion

Cuestión 8. Verificar que la función *sign* puede ser aproximada por la función *tanh*. Dado w_1 y $\epsilon > 0$ encontrar w_2 tal que $|sign(x_n^T w_1) - tanh(x_n^T w_2)| \leq \epsilon$ para $x_n \in \mathcal{D}$. Ayuda: Analizar la función $tanh(\alpha x), \alpha \in \mathbb{R}$.

Solución. solucion

Cuestión 9. Sea V y Q el número de nodos y pesos en una red neuronal,

$$V = \sum_{l=0}^{L} d^{(l)}, \quad Q = \sum_{l=1}^{L} d^{(l)} \left(d^{(l+1)} + 1\right)$$

En términos de V y Q, ¿Cuántas operaciones se realizan en un pase hacia adelante (sumas, multiplicaciones, y evaluaciones de θ)? Ayuda: Analizar la complejidad en términos de V y de Q.

Solución. solucion

Cuestión 10. Para el perceptrón sigmoidal $h(x) = tanh(x^T w)$, sea el error de ajuste $E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N}\left(tanh(x_n^T w) - y_n\right)^2$. Mostrar que

$$\nabla E_{in}(w) = \frac{2}{N}\sum_{n=1}^{N}\left(tanh(x_n^T w) - y_n\right)\left(1 - tanh(x_n^T w)^2\right)x_n$$

Si $w \to \infty$, ¿Qué le sucede al gradiente? ¿Cómo se relaciona esto con la dificultad de optimizar el perceptrón multicapa?

Solución. solucion